

UNCLASSIFIED

SAMI 8200962

A VIEW OF JOINT TEST AND EVALUATION

JUNE 1980

SYLVIA L. WALLER
SCIENTIFIC AND TECHNICAL ADVISOR ON T&E
ASSISTANT CHIEF OF STAFF, HQS USAF

Distribution Statement A
Approved for Public Release

AFSSA Collection

24-46
130 9710
Reproduced From
Best Available Copy

UNCLASSIFIED

19990729 108

SAMI 8200962

JT&E PROGRAM

Joint Test and Evaluation (JT&E) under sponsorship of the Director of Defense Test and Evaluation (DDT&E) in the Office of the Secretary of Defense has received increased emphasis in the last ten years based on recommendations of the 1970 Blue Ribbon Defense Panel Report on "Operational Test and Evaluation in the Department of Defense." The official definition of Joint Test and Evaluation as presented in DOD Directive 5000.3, is shown on the first viewgraph.

This briefing is a personal view of the unique characteristics and problems associated with JT&E programs.

JT&E PROGRAM

"AN OSD PROGRAM FOR JT&E, SPONSORED BY THE DDTE, STRUCTURED TO EVALUATE OR PROVIDE INFORMATION ON SYSTEM PERFORMANCE, TECHNICAL CONCEPTS, SYSTEM REQUIREMENTS OR IMPROVEMENTS, SYSTEMS INTEROPERABILITY, IMPROVING OR DEVELOPING TESTING METHODOLOGIES, OR FOR FORCE STRUCTURE PLANNING, DOCTRINE OR PROCEDURES.

DODD 5000.3 (ENCL 1), DEC 26, 1979

SUBJECT: TEST AND EVALUATION

JT&E -- A UNIQUE TYPE OF TESTING IN THE MILITARY SERVICES

Over the years, JT&E programs have assumed a unique set of problem areas resulting from the requirements involved in carrying out these programs. A list of these requirements are shown in this viewgraph.

The purpose of this briefing is to discuss, in turn, each of the items listed here as they have manifested themselves in actual JT&E programs over the last few years, and to challenge the analytic community to address the unique problems of JT&E resulting from these requirements in order to improve future JT&E programs and reduce the burden of JT&E on the Services.

JT&E -- A UNIQUE TYPE OF TESTING IN THE MILITARY SERVICES

REQUIRING

1. REPLICATIONS OF COMPLEX, TWO-SIDED COMBAT SITUATIONS
2. ACCURATE RECONSTRUCTION OF WHAT HAPPENED DURING EACH TRIAL
3. REALISM IN THREAT SYSTEM PERFORMANCE & ORGANIZATIONAL DOCTRINE
4. ENVIRONMENTAL CONDITIONS REPRESENTATIVE OF EXPECTED GEOGRAPHICAL AREAS OF COMBAT
5. MEASUREMENT OF THE IMPACT OF WEAPON SYSTEM PERFORMANCE ON MISSION SUCCESS--WITHOUT FIRING ANY WEAPONS
6. DEDICATED RESOURCES (SUCH AS AIRCRAFT) OVER LONG-PERIODS OF TIME
7. SATISFACTION (AND PROTECTION) OF MULTIPLE-SERVICE INTERESTS

REPLICATIONS OF COMPLEX, TWO-SIDED COMBAT SITUATIONS

As in any well-designed test program, a sufficient number of replications for each test condition are required in JT&E to assure a statistically valid sample of data for analysis of the results. However, in large-scale operational field testing of the type characterized by most of the JT&E programs to date, it is difficult to obtain a sufficient number of replications for valid statistical analysis of results without unacceptable limitations on test scope, unacceptable costs, and/or unacceptably long test periods.

As indicated in this viewgraph, even where sufficient numbers of replications are conceptually planned, it is difficult if not impossible to reproduce identical test conditions where hundreds of different players are involved over relatively long periods.

This leads to the need for innovative experimental design techniques to produce valid statistical results from very limited, very noisy data.

1. REPLICATIONS OF COMPLEX, TWO-SIDED COMBAT SITUATIONS

- o REPLICATIONS DIFFICULT TO ACHIEVE
 - LARGE NUMBER OF UNCONTROLLED VARIABLES
 - CONTINUOUSLY CHANGING ENVIRONMENT DURING EACH TRIAL, ESPECIALLY WITH REAL TIME KILL REMOVAL OF PLAYERS
 - CONTINUOUSLY CHANGING TACTICS DURING TEST (WITH FEEDBACK TO PLAYERS)
 - WITHOUT FEEDBACK, CONTINUED POOR PERFORMANCE MAY PRODUCE RESULTS NOT WORTH ANALYZING; TOO MUCH FEEDBACK LEADS TO GAMING
- o REPLICATIONS COSTLY -- LEADS TO PRE-SELECTION OF A FEW, CONTROLLED FACTORS IN FACTORIAL TEST DESIGN (NOT NECESSARILY THE MOST IMPORTANT FACTORS)
- o NEED FOR REPLICATION PRECLUDES USE OF LARGE-SCALE JCS TRAINING EXERCISES AS TEST ENVIRONMENT

QUESTION: ARE THERE EXPERIMENTAL DESIGN TECHNIQUES APPROPRIATE FOR CONTINUOUSLY CHANGING TEST CONDITIONS?

ACCURATE RECONSTRUCTION OF WHAT HAPPENED DURING EACH TRIAL

To analyze test results requires accurate reconstruction of what happened during each test trial. In large-scale field testing involving hundreds of players, instrumentation for reconstructing who did what, to whom, where and when is available today. However, such instrumentation is complex, very costly to operate and maintain and requires player units (such as aircraft) dedicated for the period of the test.

The challenge is to find new ways of accomplishing this objective of test reconstruction without the problems existent with current instrumentation/data-collection systems.

2. ACCURATE RECONSTRUCTION OF WHAT HAPPENED DURING EACH TRIAL

- o AUTOMATED TIME-SPACE-POSITION-INSTRUMENTATION (TSPI) AND ENGAGEMENT PAIRING INSTRUMENTATION FOR EVERY PLAYER
 - LARGE NUMBERS OF SEPARATE INSTRUMENTATION UNITS (HERCULEAN CHECK-OUT MAINTENANCE EFFORT)
 - LARGE 3-DIMENSIONAL GEOGRAPHICAL AREAS TO BE COVERED WITH EQUAL ACCURACY (NOT ACHIEVABLE WITH CURRENT TSPI SYSTEMS)
 - DEDICATED AIRCRAFT AND GROUND VEHICLES (WITH PRESENT TYPES OF PAIRING SYSTEMS)
 - CONFLICT BETWEEN INSTRUMENTATION OPERATIONAL CONSTRAINTS (E.G., SMOKE FOR LASER-PAIRING SYSTEMS) AND REALISTIC ENVIRONMENTS AND TACTICS
- o BACK-UP INSTRUMENTATION (VIDEO/AUDIO RECORDERS, MANUAL LOGS, OBSERVERS) TO RESOLVE ANOMALIES AND FILL AUTOMATED DATA DROPOUTS
 - LARGE MANUAL DATA-REDUCTION EFFORT (ALWAYS UNDER-ESTIMATED)
- o EARLY PRE-TEST AND ANALYSIS OF COMPLETE INSTRUMENTATION/DATA-COLLECTION DATA-REDUCTION PROCESS ESSENTIAL
 - TO IDENTIFY INTERFACE PROBLEMS, OVERLOAD CONDITIONS, DROPOUT AREAS ETC., IN TIME TO RESOLVE OR SEEK ALTERNATIVE SOLUTIONS

QUESTION: CAN WE APPLY NEW-TECHNOLOGY FOR POSSIBLE INEXPENSIVE, NON-DEDICATED PLAYER INSTRUMENTATION (E.G., GPS)?

REALISM IN THREAT-SYSTEM PERFORMANCE AND ORGANIZATIONAL DOCTRINE

An important requirement for testing (in which attrition of U.S. forces is an important measure of effectiveness) is the realism of the threat systems employed. Problems involved in achieving threat realism in JT&E are stated in this viewgraph. In many cases the validity of test results hinges on the credibility of the threat systems used during a test.

How to conduct valid large-scale two-sided tests with limited numbers of threat simulators/surrogates is a serious challenge to the conduct of meaningful JT&Es.

3. REALISM IN THREAT-SYSTEM PERFORMANCE AND ORGANIZATIONAL DOCTRINE

 - o REALISTIC RED-FORCE ESSENTIAL TO CREDIBILITY OF RESULTS
 - o FIELDED U.S. MANUFACTURED THREAT SIMULATORS USUALLY NOT REPRESENTATIVE OF LATEST THREAT SYSTEMS
 - o SURROGATING POTENTIAL FUTURE THREATS WITH AVAILABLE U.S. SYSTEMS LEADS TO RESULTS OF QUESTIONABLE RELEVANCE (E.G., I HAWK SURROGATE IN TASVAL FOR SA-8 ENGAGEMENTS)
 - SERVICE DISTRUST OF USE OF RESULTS
 - o EXTENSIVE DOCTRINAL TRAINING OF U.S. PERSONNEL REQUIRED TO REPRESENT RED FORCE TACTICS, FIRING DOCTRINE AND BEHAVIOR
 - TRAINING OF POSSIBLE NEGATIVE VALUE TO SERVICE PERSONNEL

ENVIRONMENTAL CONDITIONS REPRESENTATIVE OF EXPECTED GEOGRAPHICAL AREAS OF COMBAT

Another hairy problem associated with JT&E and with any operational field testing is the need for environmental conditions during the test that can be related to the environmental conditions expected in potential combat areas and scenarios. The difficulty of achieving this on U.S. test ranges is reflected on this chart.

4. ENVIRONMENTAL CONDITIONS REPRESENTATIVE OF EXPECTED GEOGRAPHICAL AREAS OF COMBAT

- o DIFFICULT TO FIND TEST SITES ENVIRONMENTALLY APPROPRIATE WHICH ALSO SATISFY OTHER TEST REQUIREMENTS
 - SUFFICIENT GROUND PLAYING AREA
 - ADEQUATE AIR-SPACE
 - OPERATIONAL AIRFIELDS NEAR-BY
 - ACCEPTABLE ENVIRONMENTAL IMPACT
 - FACILITIES FOR FIELDING AND MAINTAINING LARGE FORCE OF GROUND VEHICLES, TEST PERSONNEL, ADMINISTRATIVE SUPPORT, ETC.
 - ADEQUATE INSTRUMENTATION POTENTIAL
- o CONFLICT BETWEEN REALISTIC BATTLE CONDITIONS AND CURRENT INSTRUMENTATION SYSTEMS
 - SMOKEY BATTLEFIELD UNACCEPTABLE FOR LASER PAIRING (TASVAL)
 - L-BAND ECM INTERFERENCE WITH RMS TSPI SYSTEM (EWJT)
- o MOST OBVIOUS AREAS FOR TESTING (E.G., EUROPE) USUALLY NOT FEASIBLE
 - SCHEDULING DIFFICULT
 - SEVERE SAFETY CONSTRAINTS ON TACTICS
 - PLAYING AREA ON MILITARY BASES ATYPICAL OF COUNTRYSIDE
 - COST OF TRANSPORTING/OPERATING INSTRUMENTATION, TEST FORCE, ETC.

QUESTION: ARE TEST SITES IN CANADA A POTENTIAL SOLUTION TO EUROPEAN-TYPE TERRAIN/WEATHER?

MEASUREMENT OF THE IMPACT OF WEAPON-SYSTEM PERFORMANCE ON MISSION SUCCESS --
WITHOUT FIRING WEAPONS

Wherever weapon system performance (air-to-air, air-to-ground, or ground-to-air) in attriting forces is a necessary measure of effectiveness in JT&E, there is a requirement for extending from the non-destructive measures of engagement performance which can be measured in the test, to the expected ultimate outcome of each test engagement (i.e., destruction or survival) which cannot be included.

Problems associated with this requirement are shown in this viewgraph.

5. MEASUREMENT OF THE IMPACT OF WEAPON-SYSTEM PERFORMANCE ON MISSION SUCCESS --
WITHOUT FIRING WEAPONS

- o CREDIBILITY OF TEST RESULTS WEAKENED BY END-GAME SIMULATION UNCERTAINTIES
- o REQUIRES P_K -MATRIX COVERING EXPECTED ENGAGEMENT CONDITIONS OR USE OF FLY-OUT MISSILE MODELS (IN NEAR-REAL-TIME WHERE KILL-REMOVAL REQUIRED)
 - FOR P_K TABLES DIFFICULT TO ANTICIPATE IN ADVANCE THE SCOPE OF INTERACTIVE ENGAGEMENT CONDITIONS NEEDED TO CAPTURE PRINCIPAL EFFECTS
 - FOR FLY-OUT MODELS ONLY NEAR-REAL-TIME KILL REMOVAL POSSIBLE, RESULTING IN LAG IN RESPONSE
 - MORE SOPHISTICATED REAL-TIME DATA PROCESSING FOR IMMEDIATE FEEDBACK, ON ENGAGEMENT RESULTS TO PLAYERS
- o ACCURATE REPRESENTATION OF VISUAL MISSILE-LAUNCH OR GUN-FIRING CUES DIFFICULT TO ACHIEVE--IMPACT ON RESULTS UNKNOWN
- o ACCURATE REPRESENTATION OF ELECTRONIC CUES OF MISSILE-SYSTEM TRACK AND LAUNCH (RHAW SYSTEMS) DIFFICULT TO ACHIEVE WHEN SURROGATE THREAT SYSTEMS USED (E.G., USAF RHAW NOT RESPONSIVE TO I HAWK SA-8 SURROGATE IN TASVAL)

DILEMMA: NUMBER OF ENGAGEMENTS ALONE A POOR MEASURE OF SYSTEM PERFORMANCE;
HOWEVER, GREATEST CONTROVERSY ON TEST RESULTS FROM USE OF END-GAME SIMULATIONS AND/OR P_K ASSUMPTIONS

DEDICATED RESOURCES (SUCH AS AIRCRAFT, WEAPON SYSTEMS, CREWS, ETC.) OVER RELATIVELY LONG PERIODS OF TIME

Another requirement of most JT&Es conducted to date, which has created serious problems for the Services, is the need for allocating dedicated resources (such as airplanes) for long pre-test and test periods.

This need arises when special equipment or instrumentation is required on board the aircraft or ground unit, which is not already installed on most operational units.

The problems created by this situation and possible solutions are indicated in this viewgraph.

6. DEDICATED RESOURCES (SUCH AS AIRCRAFT, WEAPON SYSTEMS, CREWS, ETC.) OVER RELATIVELY LONG PERIODS OF TIME

- o AT EXPENSE OF TRAINING AND SERVICE TEST PROGRAMS
 - TRAINING LIMITED TO A FEW SELECTED CREWS (USUALLY MORE-SENIOR AIR FORCE PILOTS)
 - BENEFITS OF TRAINING WASTED, UNLESS PILOTS ASSIGNED AS INSTRUCTORS (NOT LIKELY)
- o POSSIBLE IMPACT ON READINESS OF UNITS EMPLOYED IN TEST
- o ONE ALTERNATIVE WOULD BE TO PIGGY BACK ON JCS OR SERVICE TRAINING EXERCISES (NO DEDICATED RESOURCES) -- BUT PROBABLY NOT COMPATIBLE WITH OTHER RQMTS.
- APPLICATIONS: ACCURATE RECONSTRUCTION (INSTRUMENTATION): REALISTIC THREAT ENVIRONMENT
- INNOVATIVE IDEAS NEEDED TO BETTER UTILIZE TRAINING EXERCISES FOR JT&E
- o ANOTHER ALTERNATIVE COULD BE SERIES OF SMALL TESTS IN SUPPORT OF A JOINT COMBAT STUDY (AS PROPOSED IN AIR BASE DEFENSE JT&E)
 - STUDY OBJECTIVE TO ADDRESS JT&E OBJECTIVES
 - STUDY INITIALLY IDENTIFIES CRITICAL INPUT UNCERTAINTIES THAT DRIVE ANSWERS AND CAN BE RESOLVED BY SERIES OF SMALL TESTS.
 - TEST RESULTS FOLDED BACK INTO STUDY INPUTS
 - TRAINING EXERCISES USED AS APPROPRIATE TO CONFIRM OCCURRENCE OF INTERACTION EFFECTS ASSUMED IN STUDY

SATISFACTION (AND PROTECTION) OF EACH SERVICE'S INTERESTS

The last and perhaps most important JT&E requirement addressed in this briefing is the need for JT&E programs to address important Service problems that cannot be addressed by the Services alone--and that the Services want addressed in a JT&E.

To meet this requirement, as a minimum, the criteria listed in this viewgraph must be satisfied.

Only by making JT&E a tool of the Services to achieve their own objectives, can we stop the natural tendency of reluctant participant Services to debunk JT&E as high-cost' programs of little value to them.

7. SATISFACTION (AND PROTECTION) OF EACH SERVICE'S INTERESTS

- o CLEARLY DEFINED, REALIZABLE OBJECTIVES OF IMPORTANCE TO THE SERVICES
 - ORIGINAL PURPOSE OF TASVAL UNREALIZABLE (I.E., TO REDUCE UNCERTAINTIES IN DECISIONS ON WEAPON SYSTEM ACQUISITION, FORCE STRUCTURE AND FORCE MIX)
 - ORIGINAL OBJECTIVE OF COUNTER C³ JT&E UNREALIZABLE (I.E., TO DETERMINE MILITARY WORTH OF COUNTER C³)
 - SIGNIFICANT LIMITATIONS OF TEST IDENTIFIED AND AGREED TO EARLY IN T&E PLANNING (BEFORE START OF TESTING), AND INCLUDED IN ALL EVALUATIONS OF TEST RESULTS
 - SCENARIOS
 - RELEVANCY OF TEST ENVIRONMENT TO COMBAT CONDITIONS
 - IMPORTANT CONTROLLED FACTORS NOT INCLUDED IN TEST DESIGN
 - SAFETY CONSTRAINTS ON TACTICS
 - ARTIFICIALITIES IN CUING
 - INSTRUMENTATION CONSTRAINTS ON TACTICS AND DATA RECOVERY
 - o SIGNIFICANT LIMITATIONS IN END-GAME SIMULATIONS AND/OR P_K MATRICES IDENTIFIED AND AGREED TO EARLY IN T&E PLANNING AND INCLUDED IN ALL EVALUATIONS OF TEST RESULTS

7. SATISFACTION (AND PROTECTION) OF EACH SERVICE'S INTERESTS (CONT'D)

- o A SUCCESSFUL JT&E PROGRAM REQUIRES THAT THE SERVICES RECOGNIZE IN ADVANCE THAT THE VALUE TO THEM OF THE RESULTS OUTWEIGHS BOTH THE COST IN RESOURCES, TRAINING, AND PROGRAMS SLIPPED, AND THE RISK OF RESULTS THAT CAN BE USED AGAINST SERVICE POSITIONS.

QUESTIONS: 1. HOW TO QUANTIFY THE POTENTIAL VALUE OF JT&E RESULTS TO THE SERVICES (OR TO OSD) IN ORDER TO ANALYZE IN ADVANCE THE COST/RISK ASSOCIATED WITH A PROPOSED JT&E.
2. HOW TO STOP THE CURRENT POST-TEST SYNDROME OF THE SERVICES WHICH IS TO DESTROY THE CREDIBILITY OF A JT&E PROGRAM IF THE RESULTS APPEAR TO BE DAMAGING TO THEIR INTERESTS.